

The Structure of Diversity within New World Mitochondrial DNA Haplogroups: Implications for the Prehistory of North America

Ripan S. Malhi,¹ Jason A. Eshleman,² Jonathan A. Greenberg,³ Deborah A. Weiss,² Beth A. Schultz Shook,² Frederika A. Kaestle,⁵ Joseph G. Lorenz,⁶ Brian M. Kemp,² John R. Johnson,⁷ and David Glenn Smith^{2,4}

¹Department of Human Genetics, University of Michigan, Ann Arbor; ²Department of Anthropology, ³Center for Spatial Technologies and Remote Sensing, and ⁴California Regional Primate Research Center (CRPRC), University of California, Davis; ⁵Department of Anthropology, Indiana University, Bloomington; ⁶Coriell Institute for Medical Research, Camden, NJ; and ⁷Santa Barbara Museum of Natural History, Santa Barbara, CA

The mitochondrial DNA haplogroups and hypervariable segment I (HVSI) sequences of 1,612 and 395 Native North Americans, respectively, were analyzed to identify major prehistoric population events in North America. Gene maps and spatial autocorrelation analyses suggest that populations with high frequencies of haplogroups A, B, and X experienced prehistoric population expansions in the North, Southwest, and Great Lakes region, respectively. Haplotype networks showing high levels of reticulation and high frequencies of nodal haplotypes support these results. The haplotype networks suggest the existence of additional founding lineages within haplogroups B and C; however, because of the hypervariability exhibited by the HVSI data set, similar haplotypes exhibited in Asia and America could be due to convergence rather than common ancestry. The hypervariability and reticulation preclude the use of estimates of genetic diversity within haplogroups to argue for the number of migrations to the Americas.

Introduction

Studies of mtDNA diversity among Native Americans have been used to hypothesize several different scenarios for the initial peopling of the Americas (Torroni et al. 1992, 1993; Horai et al. 1993; Merriwether et al. 1995). These studies have focused on identifying the source population(s), the number of waves of migration, and the time of entry of Native Americans into the New World and have revealed that nearly all Native Americans belong to one of five mtDNA haplogroups: A, B, C, D, or X (Schurr et al. 1990; Forster et al. 1996). A linguistically and geographically diverse set of Native American populations contains all five of these haplogroups. This pattern of haplogroup distribution for A, B, C, and D was demonstrated by Merriwether et al. (1995) and Lorenz and Smith (1996), and for haplogroup X it was demonstrated by Smith et al. (1999). All of these haplogroups were shown to exhibit similar amounts of genetic diversity (Bonatto and Salzano 1997a; Lorenz and Smith 1997) and have been identified

in prehistoric Native North American samples (Parr et al. 1996; Stone and Stoneking 1998; Carlyle et al. 2000; O'Rourke et al. 2000; Kaestle and Smith 2001; Malhi 2001). Bonatto and Salzano (1997b) demonstrated that Greenberg's (1986) proposed Eskimo-Aleut, Na-Dene, and Amerind linguistic groups (once considered to represent three different waves of migrants from Asia to the Americas) all contain haplogroup A individuals who display the nucleotide position (np) 16111 C→T transition that is not found in Asia (the exception being the Chukchi, who gained it probably as a result of back migration; see Forster et al. 1996 and Starikovskaya et al. 1998), suggesting a single New World (Beringian) origin for this mutation. These lines of evidence suggest that North America was peopled by a single wave of Asian migrants.

The source population(s) and time of entry of these colonizers is still debated. Initially, the presence of haplogroups A, B, C, and D in Mongolian populations led researchers to propose this area as the potential homeland of Native Americans (Kolman et al. 1996; Merriwether et al. 1996). Recently, this notion has been strengthened by the discovery of haplogroup X (Derenko et al. 2001) and by the distribution of Y-chromosome haplotypes 1C and 1F (Karafet et al. 1999) in the Lake Baikal/Mongolia area. Archeological evidence of human settlement of this region predates 20,000 years before present (BP) (Goebel 1999), well before humans reached eastern Siberia (Goebel 1999). The

Received September 5, 2001; accepted for publication January 10, 2002; electronically published February 13, 2002.

Address for correspondence and reprints: Dr. Ripan S. Malhi, Department of Human Genetics, 3726 Medical Science II, 1301 E. Catherine Street, University of Michigan, Ann Arbor, MI 48109-0618. E-mail: malhi@umich.edu

© 2002 by The American Society of Human Genetics. All rights reserved. 0002-9297/2002/7004-0009\$15.00

presence of these haplogroups in ancient populations of Northeast Asia would confirm this region as the homeland of North American colonizers.

The timing of initial entry into the Americas is uncertain. Through use of estimates of mtDNA diversity and rates of mtDNA evolution, a broad range of dates (11,000–43,000 years BP) have been estimated (Torroni et al. 1994; Bonatto and Salzano 1997a; Lorenz and Smith 1997; Stone and Stoneking 1998). Although researchers have recognized the need to incorporate population history in their estimates, the wide range of dates reported in the literature for the peopling of the Americas suggests that accurate models of Native American population history, accurate models of the evolution of mtDNA, and sufficient sampling of populations in the Americas have not yet emerged. For example, a large proportion of Native American linguistic diversity is found within California, yet only a few California Native American tribes and individuals have been typed for mtDNA markers (Lorenz and Smith 1996).

Several recent studies have examined mtDNA diversity within Native Americans, to address their prehistory after the initial colonization of the New World (Batista et al. 1995; Forster et al. 1996; Lorenz and Smith 1996; Kolman and Bermingham 1997; Lorenz and Smith 1997; Kaestle and Smith 2001; Malhi et al. 2001). Multiple migrations into the Americas should be characterized by distinct sets of mtDNA haplotypes and should lead to geographic substructure in their distribution and level of diversity in the New World, reflecting the different colonizations. Lorenz and Smith (1996) analyzed the mtDNA haplogroup identities of nearly 500 Native North Americans and demonstrated significant geographic variation in frequency distributions across North America. In addition, Lorenz and Smith (1996) showed that haplogroup-frequency distribution was correlated with geography and, to a lesser degree, with language. Although haplogroup-frequency distributions vary significantly across North America, regional studies of mtDNA diversity in modern populations of the Northeast (Malhi et al. 2001) and the Southwest (Malhi 2001) have confirmed a pattern of regional continuity or similarity of haplogroup-frequency distributions in these geographic areas. Regional continuity of haplogroup frequencies extends across linguistic boundaries as well as geographic space. The interregional diversity exhibited in modern populations of North America could reflect regional differences in population histories and/or separate origins of the ancestors of populations in those different geographic regions. However, the Southeast region of North America does not display regional similarities in haplogroup-frequency distributions across geographic space or linguistic boundaries, probably owing to large amounts of genetic drift caused by the relatively

intense and recent impact of European contact in this once densely populated region (Weiss 2001).

Studies of ancient mtDNA diversity in North America reveal that Native American haplogroup-frequency distributions usually exhibit temporal (as well as geographic) continuity (Carlyle et al. 2000; O'Rourke et al. 2000; Malhi 2001). However, Kaestle and Smith (2001) demonstrated that ancient Western Great Basin populations are probably not closely related to modern populations in the same region, presumably because of the population spread of Numic speakers into the Great Basin from southern California ~1,000 years BP (Bettinger and Baumhoff 1982). Analyses of modern Native American mtDNA data have also supported the hypothesized Southern Athapaskan (Torroni et al. 1992, 1993; Lorenz and Smith 1996), Algonquian (Schultz et al. 2001), and Iroquoian (Malhi et al. 2001) population movements. These studies suggested that population movements and gene flow were not negligible forces in North American prehistory, and they complicate the interpretation of the distribution and diversity of mtDNA in North America. Studies of ancient and modern populations in northeastern and southwestern North America have demonstrated that patterns of geographic similarity in haplogroup-frequency distributions predate European contact by as much as 1,500 years BP (Stone and Stoneking 1998; Carlyle et al. 2000; O'Rourke et al. 2000; Malhi et al. 2001). However, prior to this time period, the regional patterns of haplogroup-frequency distributions are uncertain.

Ward et al. (1991) demonstrated that the Nuu-Chah-Nulth possess a level of mtDNA diversity similar to that of a much larger regional and continental population, suggesting an ancient and deep common ancestry for members of this tribe. Torroni et al. (1993) similarly hypothesized that tribalization of Native Americans occurred early in prehistory. Torroni et al. (1993) concluded that the high incidence of private mtDNA polymorphisms and the limited distribution of shared mtDNA mutations in their data set support an early-tribalization model for Native Americans. Lorenz and Smith (1996) showed that intratribal genetic homogeneity is greater than intraregion genetic homogeneity, for haplogroup-frequency distributions in North America, suggesting that Native American populations experienced tribalization early in prehistory. The purpose of the present study, which uses the largest and most diverse set of samples to date, is to analyze the structure and diversity within the five Native American haplogroups and the distribution of haplogroup frequencies among Native North American populations, to investigate major postcolonization events in North America, and to estimate the time of tribalization of Native North American groups.

Material and Methods

The haplogroup identities of 1,612 Native Americans and the mtDNA sequence of the hypervariable segment I (HVSI) region (nps 16,090–16,362) of 395 of these individuals were analyzed. Populations from which haplogroup data were collected are reported in table A1 in the Appendix.

Individuals whose mtDNAs did not belong to one of the five Native American haplogroups were not included in this analysis. Although it is possible that one or more of these individuals possess previously undocumented founding Native American mtDNA types, previous studies indicate that the frequency of “other” mtDNA types is very low and that most—or all—of these result from recent admixture (Torroni et al. 1993, 1994; Huoponen et al. 1997; Smith et al. 1999). O’Rourke et al. (2000) demonstrated that most modern Native American populations surveyed to date display a pattern of regional continuity in haplogroup-frequency distribution. These patterns of haplogroup frequencies across North America suggest that a model of isolation by distance is appropriate for the analysis of such data. Using the Kriging method in the ARC/INFO software package, we interpolated haplogroup frequencies among 36 groups from across North America. Since the use of an interpolating method introduces artificial spatial autocorrelation (Sokal et al. 1999), only results that were also supported by network analyses using haplotype data were interpreted.

Haplotype networks were constructed for each of the five haplogroups, through use of three different methods. Median-joining and reduced median networks were constructed using the NETWORK 2.0 program (Bandelt et al. 1999), and polymorphic sites were weighted on the basis of relative frequency of occurrence in our sample, to correct for mutational hotspots. Haplotype networks were also constructed using the statistical parsimony method in the TCS software program (Clement et al. 2000). Haplotype networks for B, C, D, and X were constructed in TCS through use of a three-step criterion with a 95% CI. Because of the extremely high levels of reticulation observed with the three-step criterion, the haplogroup A network was constructed using a two-step criterion with a 95% CI. The approaches employed by the different network construction methods are discussed by Posada and Crandall (2001). We attempted to resolve reticulations by using predictions from coalescent theory (see Posada and Crandall 2001); however, methodologies for resolving reticulations are invalid in cases in which haplotype frequencies are strongly affected by sampling, such as in the present study. As a result, we were unable to confidently resolve most reticulations in our networks.

On the basis of the geographic origin of samples, modern haplotypes were assigned to one of five geo-

graphic categories for comparison: Northwest, Northeast, Southwest, Southeast, and Arctic. Samples with a geographic origin above 50° N latitude were assigned to the Arctic category. All other samples were assigned to a category based on their location relative to 35° N latitude and 98° W longitude (see table A1 in the Appendix). Athapaskan-speaking individuals in the Southwest, whose mtDNA identity was assigned to haplogroup A, were placed in the Northwest region category. This follows the evidence that these haplotypes reached the Southwest via a relatively recent migration (Torroni et al. 1992, 1993; Lorenz and Smith 1996; Starikovskaya et al. 1998; Smith et al. 2000; Malhi 2001). Note that ancient haplotypes were compiled from both North and South America and were not assigned to a geographic region.

Two measures of molecular diversity were calculated for the control region (CR) sequences as measures of within haplogroup diversity. Nucleotide diversity (π) was calculated as (Nei 1987)

$$\pi = \frac{\sum_{i=1}^k \sum_{j<i}^k p_i p_j d_{ij}}{L},$$

where L is the number of loci, p_i and p_j are the frequencies of the i th and j th nucleotides, and d_{ij} is the number of differences. The value of π is sensitive to haplotype frequencies and, therefore, reflects relatively recent events that influence diversity.

θ_s was calculated as (Watterson 1975)

$$\theta_s = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}},$$

where S is the number of segregating sites and n is the sample size. θ_s is calculated independently of haplotype frequencies and reflects much older influences on diversity among haplotypes than does π . All diversity calculations were performed using the ARLEQUIN software package (Schneider et al. 1997). Both estimates of diversity were calculated using the number of individuals as well as the number of haplotypes, to observe the effect of sampling on these two measures in the present study.

Results

Haplogroup-Frequency Distribution

Gene map interpolations (figs. 1A, 2A, 3A, 4A, and 5A) indicate that the frequency of haplogroup A is highest in Canada, the eastern United States, and central Mexico, whereas the frequency of haplogroup B is high-

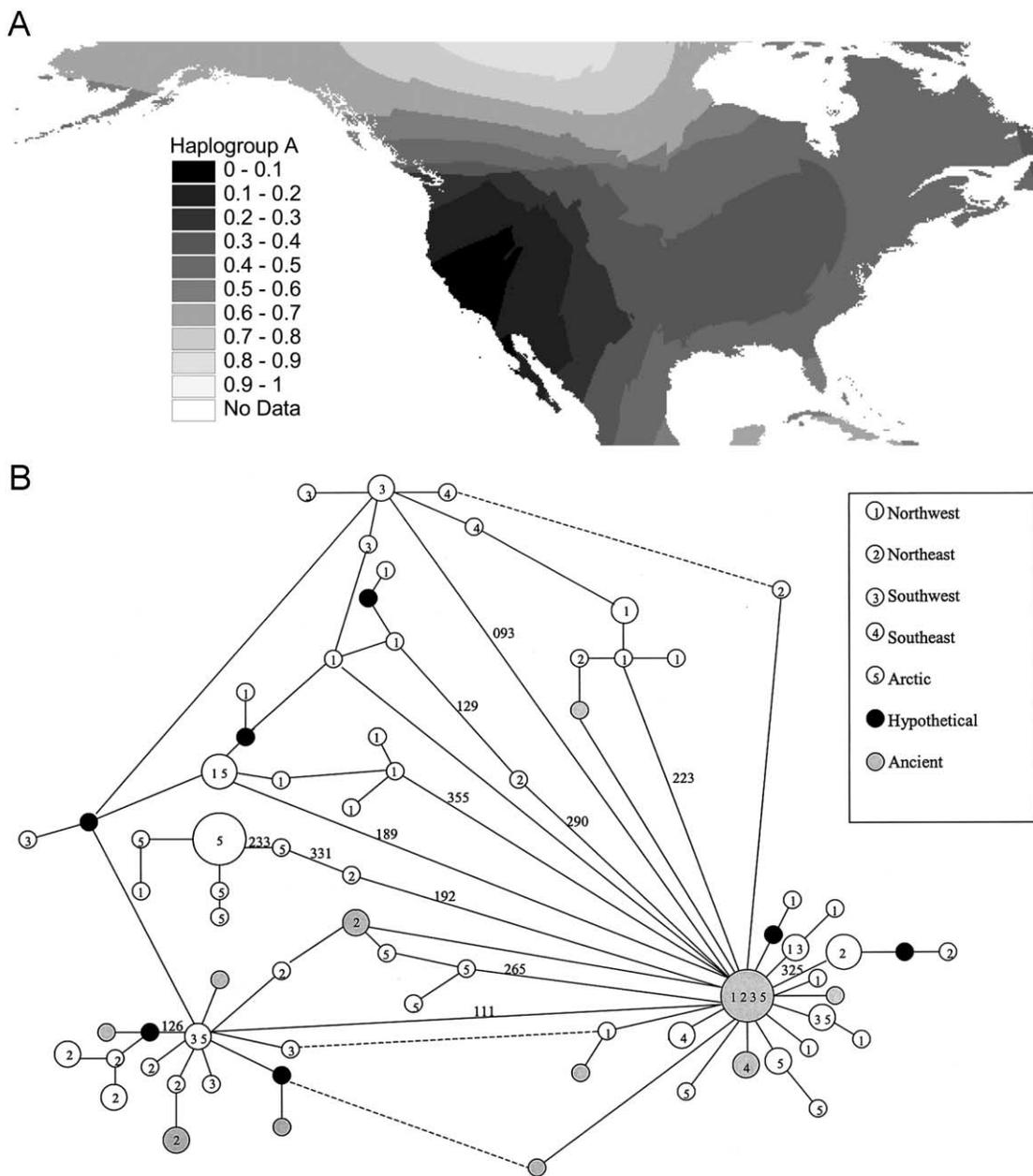


Figure 1 A, Map of haplogroup A frequencies calculated using the Kriging interpolation method. B, Two-step haplotype network for haplogroup A. The size of the circle is proportional to the frequency of that haplotype in the sample. Black circles represent hypothetical haplotypes. Each line connecting two haplotypes represents a single mutation. Mutations are indicated as the last three digits of the nucleotide position in HVSI. Dashed lines indicate possible reticulations.

est in the West and Midwest. Haplogroup C exhibits a uniform frequency throughout North America, except for a notable decrease in frequency in Alaska. Haplogroup D follows a pattern opposite that of C: frequencies are slightly higher in Alaska and lower in the remainder of North America. Haplogroup X exhibits a higher frequency around the Great Lakes and Greenland than in

the remainder of North America. The high frequency of haplogroup X in Greenland is an artifact of the interpolating methodology, since no Native American samples typed from Greenland to date can be assigned to haplogroup X (Lorenz and Smith 1996; Saillard et al. 2000). Overall, haplogroups A, B, and X exhibit strong clines.

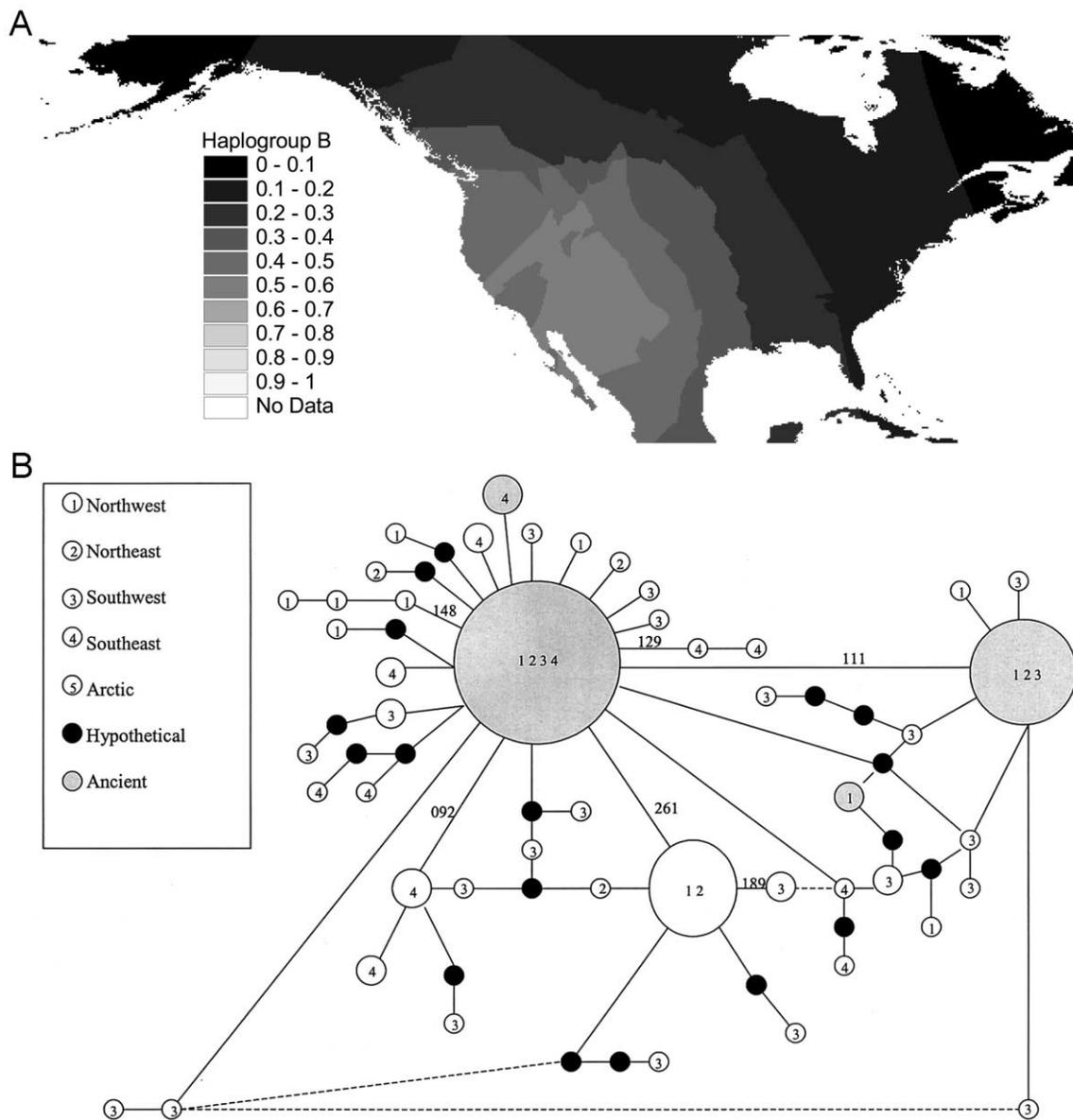


Figure 2 A, Map of haplogroup B frequencies calculated using the Kriging interpolation method. B, Three-step haplotype network for haplogroup B. For an explanation of the diagram, see figure 1.

Estimate of Molecular Diversity

Estimates of within-haplogroup diversity are given in table 1. When the diversity of unique haplotypes only is compared (an unweighted estimate), haplogroups A, B, and C exhibit similar amounts of diversity for all measures analyzed. The lower values of θ found in haplogroups D and X are probably due to the smaller number of haplotypes used in the estimate of diversity—24 and 16, respectively. When diversity measures were weighted by frequency of haplotypes, both measures of

diversity dropped. This decline was most noticeable in the calculation of π within haplogroups B and D, which appeared to have substantially lower diversity estimates under this condition.

Haplotype Distribution

An average of 29.6% of mtDNA sequences (haplotypes) are shared among Native American individuals (table 2). Ancient haplotypes were included in this estimate, even though it is possible that they are direct

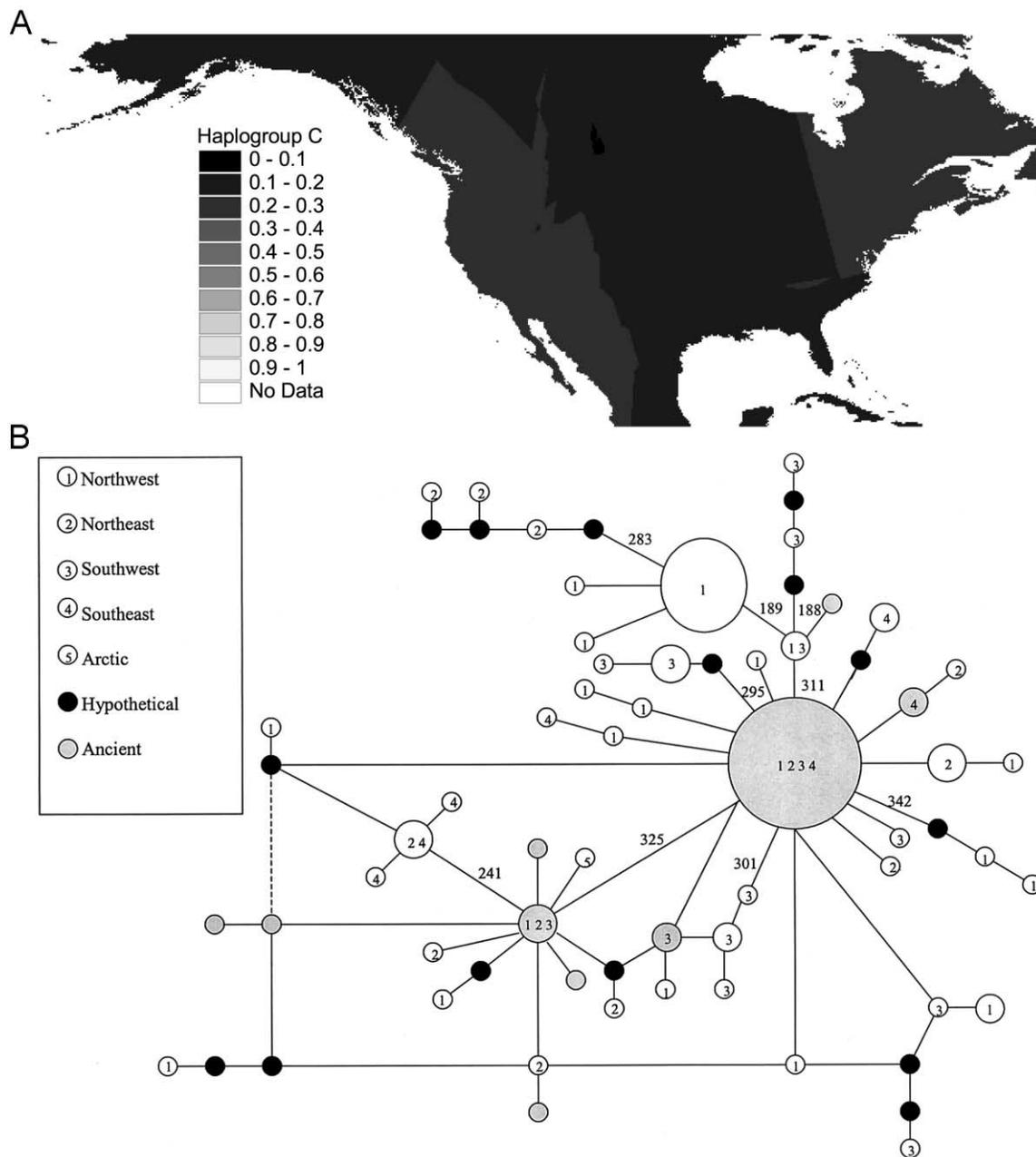


Figure 3 A, Map of haplogroup C frequencies calculated using the Kriging interpolation method. B, Three-step haplotype network for haplogroup C. For an explanation of the diagram, see figure 1.

ancestors of modern haplotypes. However, 64% of ancient haplotypes in our sample are unique and therefore have left no known descendants. Of these shared haplotypes, 40.3% are shared among geographically distant individuals, 30.8% are shared among individuals within the same region, and 28.9% are private tribal polymorphisms.

Figures 1B, 2B, 3B, 4B, and 5B illustrate the haplo-

type networks. Not surprisingly, internal nodes are generally more widely distributed geographically than external nodes, but there is no other consistent pattern among the five haplotype networks. In many instances, external haplotypes are clustered among individuals belonging to the same tribe or region, but some external haplotypes do not follow this trend. The haplogroup C network exhibits more isolation by distance than do the

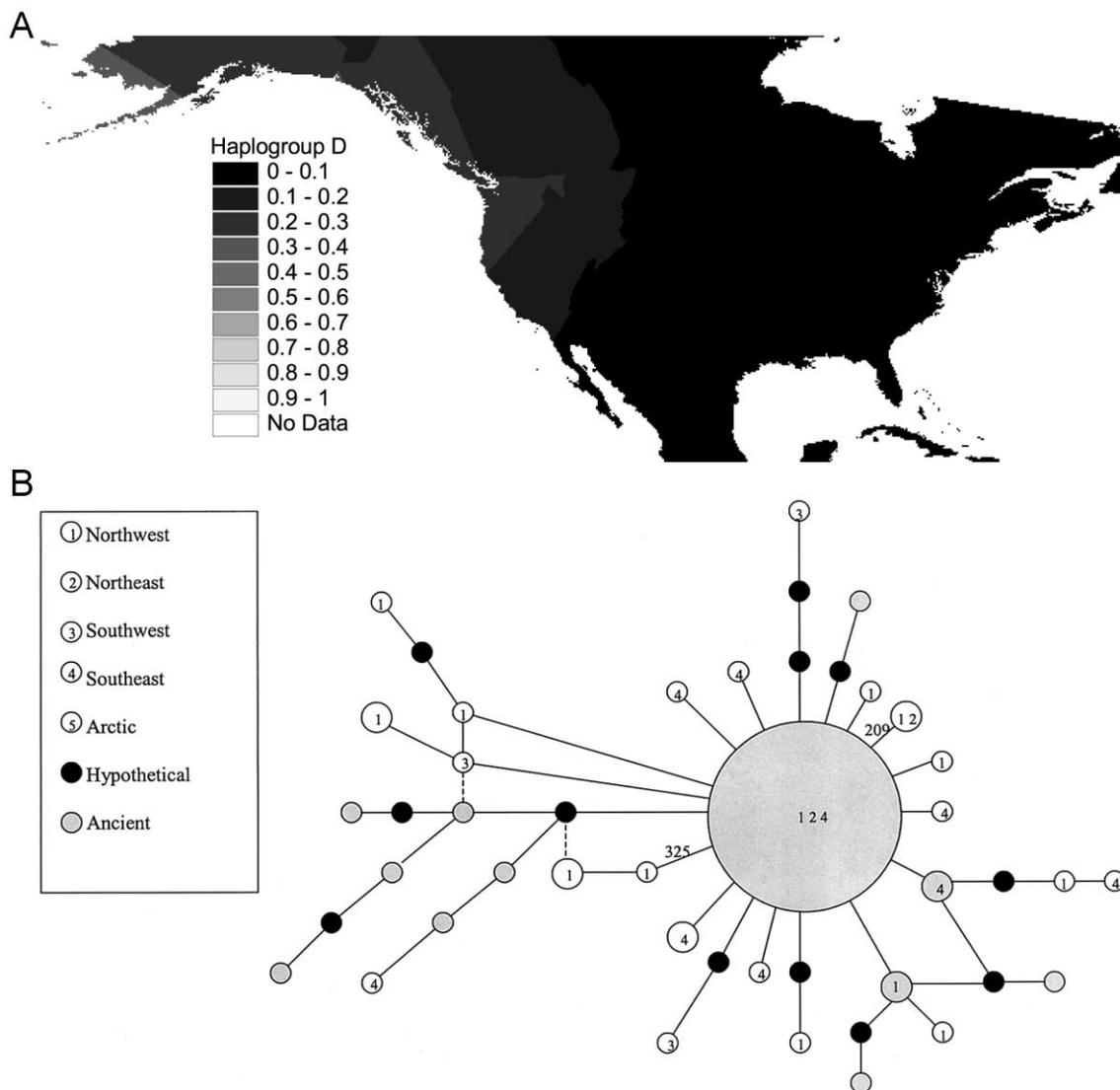


Figure 4 A, Map of haplogroup D frequencies calculated using the Kriging interpolation method. B, Three-step haplotype network for haplogroup D. For an explanation of the diagram, see figure 1.

other networks, probably because of the lack of any evidence of an expansion within that network. This network also exhibits two central haplotypes, one with a much higher frequency than the other. It is interesting to note that a Northeast clade defined by a mutation at np 16283 in that network stems from a haplotype (defined by mutations at np 16311 and np 16189) whose current distribution is limited to the Northwest. The high frequency of the haplotype defined by mutations at np 16311 and np 16189 is probably due to sampling, because this haplotype is predominantly found in one heavily sampled tribe (Northern Paiute).

The haplogroup A network is marked by numerous

reticulations and a significantly lower number of intermediate hypothetical haplotypes. As demonstrated by Malhi (2001), the haplogroup B network contains three main haplotypes, all of which are shared among two or more geographic regions. In addition to the founding haplotype defined by Forster et al. (1996), one of the main haplotypes is defined by a mutation at np 16111, and the other haplotype is defined by a mutation at np 16261. A large number of individuals in the haplogroup D network were assigned to the basal haplotype. Interestingly, this network exhibits two haplotypes, found among ancient individuals, that are internal to a modern Southeast haplotype, but most ancient haplotypes are

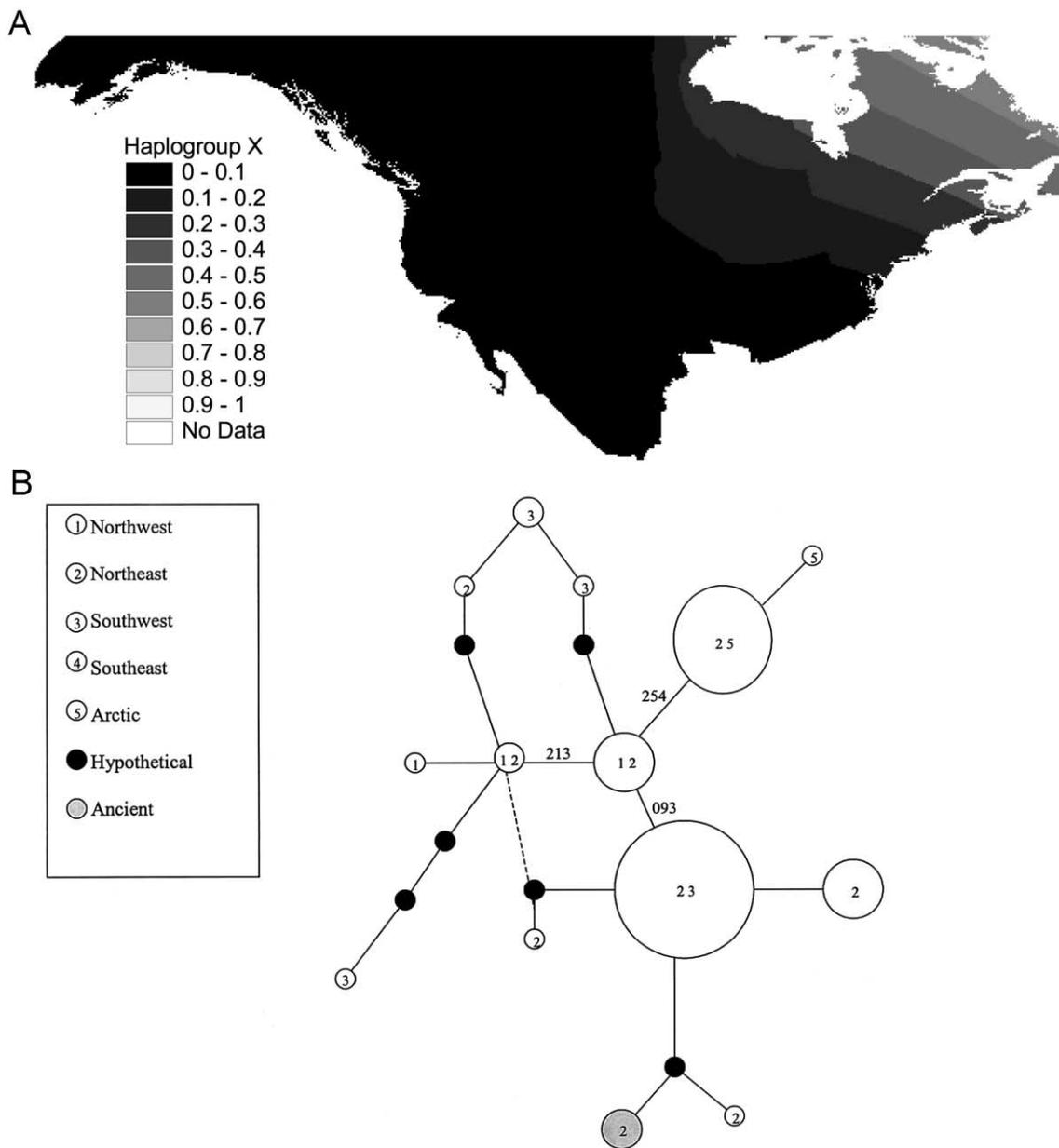


Figure 5 A, Map of haplogroup X frequencies calculated using the Kriging interpolation method. B, Three-step haplotype network for haplogroup X. For an explanation of the diagram, see figure 1.

terminal and might be extinct. The haplogroup X network exhibits haplotypes from four of the five geographic regions, but Algonquian-speaking individuals predominate in the network. The extent to which sampling contributes to this pattern is not clear.

In the present study, 36% of haplotypes from ancient samples are shared with modern Native Americans. If they are not shared, most ancient haplotypes cluster with the modern haplotypes, suggesting that European con-

tact did not cause a significant change in haplotype structure in most regions of North America. In the Southeast, however, four different ancient haplotypes of haplogroup D form intermediate nodes that connect highly divergent modern external haplotypes with internal haplotypes found in ancient Native Americans. This suggests that, unlike other regions of North America, the Southeast experienced a high percentage of haplotype extinctions. This pattern is consistent with the random distri-

Table 1**Diversity Estimates**

Haplogroup	N	π	θ_s	Sequence Divergence ^a
Unweighted:				
X	16	.183	5.123	1.201
A	60	.079	9.650	1.341
B	47	.071	10.415	1.252
C	45	.088	9.834	1.397
D	24	.089	8.569	1.201
Weighted:				
X	46	.143	3.868	.935
A	83	.072	9.018	1.215
B	80	.052	9.287	.910
C	75	.067	8.797	1.067
D	54	.048	7.242	.661

NOTE.—Ancient haplotypes are not included in the estimates.

^a Based on 290 base pairs.

bution of haplogroup frequencies among closely related populations from the Southeast, which may indicate a recent bottleneck as a result of European contact (Weiss 2001).

Discussion

Hypervariable Sites and the Control Region

Analysis of the median joining, reduced median, and statistical parsimony networks for the 290-bp segment of HVSI revealed significant reticulation in all haplogroups that was not readily resolved using a coalescence method. Although the future discovery of new haplotypes might resolve some of these reticulations, the high degree of reticulation could be the result of the hypervariable nature of this region, since a number of sites have previously been reported to be hypermutable (Hasegawa et al. 1993; Wakeley 1993; Excoffier and Young 1999; Meyer et al. 1999; Gurven 2000; Sigurðardóttir et al. 2000; Stoneking 2000). Many sites appear to have mutated in multiple haplogroups; for example, the haplotype networks for haplogroups A and B both exhibit mutations at np 16111, np 16129, and np 16189. However, hypervariability need not be implicated in all instances of multiple hits at a nucleotide position. In the most conservative assessment of mutational positions, where mutations are distributed entirely at random throughout the 290-bp region, the probability that each mutational site among n such sequences is unique is $290 \text{ Permut } n/(290^n)$; with 21 variable positions among the 290 bases analyzed, the probability of ≥ 1 position having mutated more than once is $>50\%$, even if all sites are equally likely to change. Since haplotypes within haplogroups A, B, C, and D exhibited 45, 45, 43, and

27 polymorphic sites, respectively, the occurrence of the same mutational position in different haplogroups is not unlikely, especially in the case of the three mutations cited above that are shared by members of haplogroups A and B. Likewise, some reticulation within a haplotype network, caused by a second occurrence of the same mutation within a separate sublineage, should be expected even in the complete absence of any hypervariable sites. Only within the network for haplogroup X was there a $<50\%$ chance that a single mutated position would occur in separate haplotypes without any hypervariable sites. Even in this instance, where the haplogroup exhibited 17 variable positions with equal probability of mutating, there is only a 62% chance that a mutation has occurred in only one lineage.

The high rate of population growth of the migrants who colonized the Americas and of their descendants might be responsible for the high number of reticulations exhibited by Native American haplotype networks. Higher rates of reticulation and obscured ancestry will be observed in haplogroups that have experienced population expansions, since the potential for obscuring ancient ancestry increases with the increased numbers of mutational events experienced in a population expansion. Therefore, haplogroups C and D, which showed less reticulation in all haplotype networks generated, should be more reliable for reconstructing distant shared ancestry among maternal lineages in North America than haplogroups A, B, and X.

Paleo-Archaic Population History

Multiple lines of evidence (discussed in the Introduction) suggest that Native Americans descend from migrants of a single source population that colonized the Americas in a single wave. Under the assumption of a single migration and source population for colonizers of the Americas, the Northeast clade that is defined by a mutation at np 16283 and that stems from the Northwest clade in the haplogroup C network suggests a west-to-east migration of Native Americans during prehistoric times. The large amount of variation in the distribution of haplogroup frequencies among geographic regions suggests that early colonizers spread across the continent in small bands that were strongly affected by genetic drift. Nevertheless, the fact that 40% of shared haplotypes among Native Americans from North America are shared among geographically distant individuals suggests that the early inhabitants of this region experienced substantial amounts of gene flow.

Long-distance migration cannot explain the distribution of shared haplotypes, because many haplotypes are shared by geographically distant individuals who speak unrelated languages, unlike the pattern usually observed in genetic analysis of prehistoric migrations

Table 2**Distribution of Haplotypes within a Haplogroup**

HAPLOGROUP (N)	PERCENTAGE				
	Ancient Shared with Modern ^a	Basal Haplotype	Shared Haplotypes	Component of Shared Haplotypes ^b	
				Tribal Shared Polymorphisms	Distant Shared Polymorphisms
A (70)	38.46	7.22	24.29	17.65	35.29
B (56)	66.67	20.21	21.43	33.33	33.33
C (51)	45.45	16.67	23.53	25.00	50.00
D (35)	28.57	36.51	20.00	28.57	42.86
X (17)	1.00	3.51	58.82	40.00	40.00
Average	36.03	16.82	29.61	28.91	40.30
SD	24.06	12.93	16.41	8.44	6.60

^a Indicates the percentage of nonunique ancient haplotypes.

^b Includes ancient and modern samples.

(Lorenz and Smith 1996; Kaestle and Smith 2001; Malhi et al. 2001). It is possible that tribalization followed by a significant reduction of gene flow occurred very early in North American prehistory and that the high percentage of shared distant haplotypes is due to convergence caused by the high rate of mutation in the HVSI region. Alternatively, early Native American populations might have dispersed into small bands that were sufficiently mobile to maintain significant levels of gene flow while being small enough to be strongly affected by genetic drift. The Multiregional Continuity hypothesis for the spread of *Homo sapiens sapiens* worldwide employs this model to explain the maintenance of regional genetic diversity within a single lineage experiencing rapid population expansion (Templeton, in press). This scenario fits well with the interpretation of the early archeological record of North America that portrays the continent's earliest inhabitants as highly mobile big-game hunters who used a fluted-tip technology to follow game throughout a large home range (Kelly and Todd 1988).

After the onset of the Archaic period, prehistoric Native North Americans in different regions began to specialize and intensify methods to procure local resources, at different time periods. For example, the trend toward resource specialization occurs in the Northwest, where large coastal shell mounds, large fishing sites, and pit-houses on the Columbia River appeared ~6,000 years BP. In other regions of North America, such as the Southwest, intensification of resource utilization resulted in the development of agriculture ~3,500–1,500 years BP (Fagan 2000). The appearance of resource intensification in the archeological record probably reflects a time of increased population growth rates in a region. Intensification allowed hunter-gatherer groups to become much more sedentary than their ancestors. The resulting increase in population size reduced the influence of genetic drift within a group while also focusing gene flow in

local groups within the same geographic region. Therefore, the time period of intensification probably marks the beginnings of the pattern of strong regional continuity observed in haplogroup-frequency distributions in Native North Americans today.

Although the unweighted haplotype diversity estimates for the five Native American haplogroups are similar, the haplogroup-frequency maps and the structure of diversity within the haplotype networks suggest that the five Native American haplogroups experienced significantly different population histories after the colonization of the Americas. Specifically, North American populations in regions with high frequencies of haplogroups A, B, and X, respectively, all appear to have undergone population expansions, whereas regions containing populations with high frequencies of haplogroups C and D exhibit little or no evidence of population expansions. The clinal distribution of haplogroup A in North America suggests that populations with high frequencies of this haplogroup experienced a population expansion in the northern regions of North America. The haplogroup A network supports this trend, by exhibiting a large amount of structure and a much higher frequency of nonbasal haplotypes in the north than in the south. In addition, all haplotypes that were too divergent to fit in a two-step clade were found in "Amerindian-speaking" individuals from the northern regions of North America, suggesting that the greatest amount of diversity occurs in the northern area. Starikovskaya et al. (1998) demonstrated that this expansion occurred in a clade of haplogroup A defined by np 16192T, and Shields et al. (1993) and Forster et al. (1996) reported that populations located in the northern regions of North America (Na-Dene and Eskimo speakers) contain about half the total extant diversity within haplogroup A in North America. Thus, this expansion led to the replacement of older non-A native haplotypes in the north.

The gene map suggests that haplogroup B experienced an expansion in the southwestern region of North America. This expansion broadened the distribution and increased the frequency of the haplogroup B clade, particularly one subclade of B in the Southwest. Defined by a T→C mutation at np 16261 (Malhi 2001), this subclade comprises 82% of haplogroup B mtDNA in North America. The high values of θ_s for haplogroups B and C, which predominate in the Southwest (Malhi 2001), also suggest an early population expansion in this region. There is archeological evidence of a population expansion in the Southwest during early Holocene times. The oldest Clovis-associated dates, ~11,550 radiocarbon years BP (rcbp) (13,400 years BP), come from Texas (the Aubrey site; Fiedel 1999) and Clovis sites that subsequently appear throughout North America and that undergo a stylistic transition in Central America into Fell's Cave fishtail points. These fishtail-type points are then carried throughout South America, reaching Tierra del Fuego 11,000 rcbp (Fiedel 1999).

Additional evidence of an early population expansion in the Southwest is provided by Fisher et al. (2001), whose phylogenetic analysis of microsatellite data demonstrated that the common ancestor for all variants of Valley Fever (*Coccidioides immitis*) in South America is located in the American Southwest. They suggest that Valley Fever spread from the American Southwest to Central and South America, some time before 9,000 years BP, as a commensal of humans. A much later population expansion associated with the development and use of agriculture in the Southwest, ~3,500–1,500 years BP (Fagan 2000), probably contributed the remainder of the variation in this clade of haplogroup B.

Our sample of haplogroup X consists of a large percentage of shared haplotypes among tribes speaking Chippewa/Ojibwa languages and dialects. The haplogroup X network and distribution of haplogroup frequencies suggest that populations with relatively high frequencies of haplogroup X experienced an expansion in the Great Lakes region. This expansion, which generated a value of θ_s only half that for haplogroup B, must have occurred much more recently in prehistory than the expansion of haplogroup B. Archeological, linguistic, and genetic evidence all strongly support the expansion of Algonquian-speakers from the Great Lakes region, ~2,500–3,000 years BP (Denny 1991; Malhi et al. 2001). Ancient-DNA studies of prehistoric populations from the Great Lakes region demonstrated that this Algonquian expansion probably occurred 700–3,000 years BP (Schultz et al. 2001).

Although the marked variation among haplogroups in the weighted estimates of π could be interpreted as evidence for multiple waves of colonization, these estimates are highly sensitive to sampling error. Specifically, a high percentage of the haplogroup B sequences is lim-

ited to the Southwest region, and many of these are basal sequences. Although this could arise from a more recent entry into the Southwest, as some have argued (Torroni et al. 1992), a recent re-expansion in this region could also result in an overrepresentation of basal lineages in the weighted sample. Indeed, under this condition, π would not reflect the diversity accumulated since colonization from Asia but rather a recent expansion resulting from the introduction of agriculture to this region within the last few thousand years. Values for θ_s , a statistic that reflects long-term diversity, were much more similar among haplogroups (with the single exception of haplogroup X, where sampling is biased in favor of Algonquian-speaking populations) than were values of π , lending support to the position that the five haplogroups have roughly the same level of accumulated diversity.

Additional Founding Haplotypes

Since individuals with haplotype B exhibiting the np 16261C mutation are also found in Mongolia (Kolman et al. 1996), south central China (Yao et al. 2000), South America (Ward et al. 1996) and western North America, this haplotype might be an additional founding haplotype of the Americas. However most of these haplotypes are restricted to the Southwest, with only a single individual exhibiting this mutation in the Northwest in North America. The limited geographic distribution of haplotype B with the np 16261C mutation suggests that similarities between this haplotype in Asia and America might result from convergence. The haplogroup B network also includes many geographically diverse samples exhibiting a mutation at np 16111C, which is also found in haplogroup B individuals in south central China (Yao et al. 2000). However, since np 16111 is generally regarded as a hypervariable site in Native American populations, we cannot currently demonstrate that haplotype B with the np 16111C mutation is an additional founding haplotype. The co-occurrence of this haplotype in Asia and the Americas might also be due to convergence.

Members of haplogroup C who exhibit the np 16325C mutation are found in both Asia (Torroni et al. 1993) and the Americas, with members in the latter area being widely distributed throughout many regions of North America. However, since the mutation at np 16325, like that at np 16111, might also be hypervariable, we cannot confidently regard haplotype C with the np 16325C mutation as a founding haplotype; its presence in both Asia and the Americas could result either from a migration of a single type with np 16325T and subsequent independent C→T transitions in both the New and Old World or from the migration of two founding lineages, one with and one without the mutation at np 16325.

Brown et al. (1998) demonstrated that Europeans as-

signed to haplogroup X lack a mutation at np 16213 in the HVSI that all Native Americans exhibit. However, the larger sample size of individuals assigned to haplogroup X in the present study reveals that a substantial number of Native Americans in multiple geographic regions also lack the np 16213G mutation and therefore have haplotypes identical to those of European (Brown et al. 1998) and Asian (Derenko et al. 2001) members of haplogroup X. A central X haplotype is shared among Native Americans in the Northwest and Northeast, suggesting that this haplotype might be the founding X haplotype in eastern North America. Smith et al. (1999) demonstrated that haplogroup X is present in a more linguistically diverse population in the Northwest, whereas in the Northeast this haplogroup is mainly limited to Algonquian speakers. This is consistent with the hypothesis that haplogroup X was first introduced to the eastern part of North America by Algonquians emigrating from northwestern North America (Malhi et al. 2001; Schultz et al. 2001).

The present study raises doubt about interpretations of previously reported evidence for the number of migrations to the Americas. If substantiated, the presence of additional founding haplotypes within haplogroups B and C in the New World would significantly reduce previous estimates of diversity accumulated since colonization within these haplogroups. Many researchers (Bonatto and Salzano 1997a; Lorenz and Smith 1997; Brown et al. 1998; Stone and Stoneking 1998) have interpreted similar estimates among at least four of the five Native American haplogroups as evidence that all haplogroups entered the Americas at the same time. In contrast, Torroni et al. (1994) employed exhaustive restriction analysis to argue that the lower diversity within haplogroup B suggests a later migration of this haplogroup to the New World. These estimates rely, at least in part, on knowing the level of diversity of haplotypes at the time of initial colonization. If multiple haplotypes within a haplogroup were successful colonizers of the New World, modern values of within-haplogroup diversity would overestimate the accumulated variation since contact.

In addition, the HVSI portion of the control region used to create estimates of genetic diversity in many studies exhibits a high percentage of polymorphic sites, suggesting that the entire region itself is hypervariable. If so, many nucleotide sites will experience multiple hits, resulting in back mutations, and certain mutational sites will mutate independently in separate lineages. Comparisons of mutational sites between haplogroups and of mutational sites within haplotype networks strongly suggest that both events have occurred. The level of diversity that accumulates after this saturation point has

been reached will result in a nonlinear accumulation of mutations within a haplogroup, thereby further impairing the utility of molecular diversity for the dating of the colonization event. Diversity estimates are also strongly affected by both sampling and population historic and demographic events that have occurred since colonization. Previously reported lower diversity estimates for haplogroup B may well be a result of the more recent expansion of this haplogroup within the Southwest, an area that has been well represented—and sometimes overrepresented—in many studies of Native American mtDNA diversity. The lower levels of diversity within haplogroup B might actually be reflecting this expansion rather than an earlier colonization of the Americas. In this light, we believe that the wide distribution of haplogroups throughout North America is strong evidence for a single entry from Asia.

The high rate of mutation of HVSI suggests that control region data are useful for identifying diagnostic mutations that are specific to a tribe. However, older mutations that are potentially regionally specific or shared among all Native Americans are obscured by the high frequency of multiple hits at nucleotide positions in this region. Therefore, we are unable to distinguish whether shared mutations among geographically distant individuals are due to prehistoric Native Americans existing as one population with little substructure for an extended period of time prior to intensification of resource utilization during Archaic times or, alternatively, due to convergence. Analysis of polymorphic sites in a less mutable region of the mitochondrial genome, such as a coding region, in addition to the sites in the control region, might help resolve some of the reticulations in Native American haplotype networks, as was done with European haplotype networks (Finnila and Majamaa 2001). This would allow us to better date the time of tribalization of prehistoric Native American groups and to identify additional founding haplotypes.

Acknowledgments

We would like to thank Dr. Robert Bettinger and two anonymous reviewers, for their suggestions and insights. We are indebted to numerous personnel of Indian Health Service Facilities, where most of the samples studied were obtained, as well as to individuals who provided samples used in this analysis and to the Native Americans who authorized their use. This study was supported by National Institutes of Health grants RR00169 and RR05090, by National Science Foundation grants GER9255683 and SBR9630926, and by a National Science Foundation dissertation improvement grant (to R.S.M.).

Appendix

Table A1

Populations from Which Haplogroup Data Were Collected

Population	Sample Size	Geographic Location	Reference for RFLP Data
Navajo	64	Southwest/haplogroup A: Northwest	Malhi 2001
Northern Paiute	98	Northwest	Kaestle and Smith 2001
Cheyenne/Arapaho	35	Northeast	Malhi et al. 2001
Washo	38	Northwest	Lorenz and Smith 1996
Yokut	17	Southwest	Lorenz and Smith 1996
Havasupai	18	Southwest	Lorenz and Smith 1996
Quechan	23	Southwest	Lorenz and Smith 1996
Kumeyaay	16	Southwest	Lorenz and Smith 1996
Apache	38	Southwest/haplogroup A: Northwest	Malhi 2001
Pima	43	Southwest	Malhi 2001
Hopi	4	Southwest	Lorenz and Smith 1996
Sioux	45	Northeast	Malhi 2001
Mohawk	123	Northeast	Merriwether and Ferrell 1996
Ojibwa	33	Northeast	Scozzari et al. 1997
T.M. Chippewa	28	Northeast	Malhi et al. 2001
Pawnee	5	Southeast	Malhi et al. 2001
Stillwell Cherokee	37	Southeast	Malhi et al. 2001
Zuni	26	Southwest	Malhi et al. 2001
Jemez	36	Southwest	Malhi 2001
Eskimo	115	Arctic	Merriwether et al. 1995
Aleut	72	Arctic	Merriwether et al. 1995
Creek	35	Southeast	Weiss 2001
Choctaw	27	Southeast	Weiss 2001
Yakima	42	Northeast	Shields et al. 1993
Micmac	6	Northeast	Malhi et al. 2001
Northern Hohan	6	Northwest	Lorenz and Smith 1996
Dogrib	42	Arctic	Merriwether et al. 1995
Bella Coola	36	Northwest	Lorenz and Smith 1996
Wishram	20	Northwest	Malhi 2001
CA Uto-Aztecan	14	Southwest	Lorenz and Smith 1996
Kiliwa	7	Southwest	Malhi 2001
North Central Mexico	199	Southwest	Green et al. 2000
Seminole	35	Southeast	Huoponen et al. 1997
Greenland Eskimo	82	Arctic	Saillard et al. 2000
Chumash	21	Southwest	Lorenz and Smith 1996
Nahua	31	Southwest	Malhi 2001

NOTE.—Modern HVSI sequence data were obtained from Ward et al. (1991, 1993), Torroni et al. (1992, 1993), Shields et al. (1993), Lorenz and Smith (1997), Kaestle (1998), Malhi et al. (2001), Weiss (2001), and J.G.L. (unpublished data). Ancient HVSI sequence data were taken from Lalueza-Fox (1996), Ribeiros dos Santos et al. (1996), Kaestle (1998), Stone and Stoneking (1998), and Malhi (2001).

References

- Bandelt HJ, Forster P, Rohlf A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Batista O, Kolman CJ, Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panama. *Hum Mol Genet* 4:921–929
- Bettinger RL, Baumhoff MA (1982) The Numic spread: Great Basin cultures in competition. *Am Antiq* 47:485–503
- Bonato SL, Salzano FM (1997a) Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am J Hum Genet* 61:1413–1423
- (1997b) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 94:1866–1871
- Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC (1998) mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am J Hum Genet* 63:1852–1861
- Carlyle SW, Parr RL, Hayes MG, O'Rourke, DH (2000) Context of maternal lineages in the greater Southwest. *Am J Phys Anthropol* 113:85–101
- Clement M, Posada D, and Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659

- Denny P (1991) The Algonquian migration from Plateau to Midwest: linguistics and archaeology. In: Cowan W (ed) Papers of the 22nd Algonquian Conference. Carleton University Press, Ottawa, pp 86-107
- Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miścicka-Śliwka D, Zakharov IA (2001) The presence of mitochondrial haplogroup X in Altaians from South Siberia. *Am J Hum Genet* 69:237-241
- Excoffier L, Young Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Mol Biol Evol* 16:1357-1368
- Fagan BM (2000) Ancient North America: the archaeology of a continent. Thames & Hudson, New York
- Fiedel SJ (1999) Older than we thought: implications of corrected dates for Paleoindians. *Am Antiq* 64:95-115
- Finnila S, Majamaa K (2001) Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population. *J Hum Genet* 46:64-69
- Fisher MC, Koenig GL, White TJ, San-Blas G, Negroni R, Alvarez IG, Wanke B, Taylor JW (2001) Biogeographic range expansion into South America by *Coccidioides immitis* mirrors New World patterns of human migration. *Proc Natl Acad Sci USA* 98:4558-4562
- Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of native American mDNA variation: a reappraisal. *Am J Hum Genet* 59:935-945
- Goebel T (1999) Pleistocene human colonization of Siberia and peopling of the Americas: an ecological approach. *Evol Anthropol* 8:208-227
- Green LD, Derr JN, Knight A (2000) mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* 66:989-998
- Greenberg J, Turner CG II, Zegura SL (1986) The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Curr Anthropol* 4:477-497
- Gurven M (2000) How can we distinguish between mutational "hot spots" and "old sites" in human mtDNA samples? *Hum Biol* 72:455-471
- Hasegawa M, Rienzo AD, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* 37:347-354
- Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* 10:23-47
- Huoponen K, Torroni A, Wickman PR, Sellitto D, Gurley DS, Scozzari R, Wallace DC (1997) Mitochondrial DNA and Y chromosome-specific polymorphisms in the Seminole of South Florida. *Eur J Hum Genet* 5:25-34
- Kaestle FA (1998) Molecular evidence for prehistoric Native American population movement: the Numic expansion. PhD thesis, University of California, Davis
- Kaestle FA, Smith DG (2001) Ancient mitochondrial DNA evidence for prehistoric population movement: the Numic expansion. *Am J Phys Anthropol* 115:1-12
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817-831
- Kelly RL, Todd LC (1988) Coming into the country: early Paleoindian hunting and mobility. *Am Antiq* 53:231-244
- Kolman CJ, Bermingham E (1997) Mitochondrial and nuclear DNA diversity in the Choco and Chibcha amerinds of Panama. *Genetics* 147:1289-1302
- Kolman CJ, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321-1334
- Lalueza-Fox C (1996) Analysis of ancient mitochondrial DNA from extinct aborigines from Tierra del Fuego/Patagonia. *Anc Biomol* 1:43-55
- Lorenz JG, Smith DG (1996) Distribution of four founding mtDNA haplogroups among Native North Americans. *Am J Phys Anthropol* 101:307-323
- (1997) Distribution of sequence variation in the mtDNA control region of native North Americans. *Hum Biol* 69:749-776
- Malhi RS (2001) Investigating prehistoric population movements in North America using ancient and modern mtDNA. PhD thesis, University of California, Davis
- Malhi RS, Schultz BA, Smith DG (2001) Distribution of mitochondrial DNA lineages among native American tribes of northeastern North America. *Hum Biol* 73:17-55
- Merriwether DA, Ferrell RE (1996) The four founding lineage hypothesis for the New World: a critical reevaluation. *Mol Phylogenet Evol* 5:241-246
- Merriwether DA, Hall WW, Vahlne A, Ferrell RE (1996) mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. *Am J Hum Genet* 59:204-212
- Merriwether DA, Rothhammer F, Ferrell RE (1995) Distribution of the four founding lineage haplotypes in native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol* 98:411-430
- Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103-1110
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- O'Rourke DH, Hayes MG, Carlyle SW (2000) Spatial and temporal stability of mtDNA haplogroup frequencies in native North America. *Hum Biol* 72:15-34
- Parr RL, Carlyle SW, O'Rourke D (1996) Ancient DNA analysis of Fremont Amerindians of the Great Salt Lake wetlands. *Am J Phys Anthropol* 99:507-518
- Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* 16:37-45
- Riberos dos Santos AK, Santos SE, Machado AL, Guapindaia V, Zago MA (1996) Heterogeneity of mitochondrial DNA haplotypes in Pre-Columbian natives of the Amazon region. *Am J Phys Anthropol* 101:29-37
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718-726
- Schneider S, Roessli D, Excoffier DL (1997) Arlequin: a software for population genetic data analysis. University of Geneva, Geneva
- Schultz BA, Malhi RS, Smith DG (2001) Examining the Proto-Algonquian migration: analysis of mtDNA. In: Nichols JD,

- Ogg A (eds) Proceedings of the 32d Algonquian Conference. Carleton University Press, Ottawa, pp 470–492
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 46:613–623
- Scozzari R, Cruciani F, Santolamazza P, Sellitto D, Cole DE, Rubin LA, Labuda D, Marini E, Sacca V, Vona G, Torroni A (1997) mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. *Am J Hum Genet* 60:241–244
- Shields GF, Schmeichen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet* 53:549–562
- Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66:1599–1609
- Smith DG, Lorenz J, Rolfs BK, Bettinger RL, Green B, Eshleman J, Schultz B, Malhi R (2000) Implications of the distribution of Albumin Naskapi and Albumin Mexico for new world prehistory. *Am J Phys Anthropol* 111:557–572
- Smith DG, Malhi RS, Eshleman J, Lorenz JG, Kaestle FA (1999) Distribution of mtDNA haplogroup X among Native North Americans. *Am J Phys Anthropol* 110:271–284
- Sokal, RR, Oden, NL, Thomson, BA (1999) Problems with synthetic maps remain: reply to Rendine et al. *Hum Biol* 71: 447–453
- Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am J Hum Genet* 63:1473–1491
- Stone AC, Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet* 62:1153–1170
- Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet* 67: 1029–1032
- Templeton AR. Out of Africa again and again. *Nature* (in press)
- Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) Mitochondrial DNA “clock” for the Amerinds and its implications for timing their entry into North America. *Proc Natl Acad Sci USA* 91:1158–1162
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53:563–590
- Torroni A, Schurr TG, Yang CC, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162
- Wakeley J (1993) Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. *J Mol Evol* 37:613–623
- Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 88:8720–8724
- Ward RH, Redd A, Valencia D, Frazier B, Pääbo S (1993) Genetic and linguistic differentiation in the Americas. *Proc Natl Acad Sci USA* 90:10663–10667
- Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV (1996) Mitochondrial DNA polymorphism in three Brazilian Indian tribes. *Am J Hum Biol* 8:317–323
- Watterson G (1975) On the number of segregation sites in the genetical models without recombination. *Theor Popul Biol* 7:256–276
- Weiss DA (2001) Mitochondrial DNA diversity among Native Americans from the southeastern United States. *Am J Phys Anthropol Suppl* 32:163
- Yao YG, Watkins WS, Zhang YP (2000) Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of east and southeast Asia. *Hum Genet* 107:504–512